

Processamento de Fala 2011/12

3º Mini-teste

26 de Maio de 2012

Identifique o seu teste colocando o seu nome e número de aluno no espaço reservado na última página. Só são aceites respostas às questões de escolha múltipla assinaladas no local apropriado nessa página. As respostas erradas a perguntas de escolha múltipla serão penalizadas. Use o alfabeto fonético SAMPA (região de Lisboa).

1. Diga se as seguintes afirmações são verdadeiras ou falsas:
 - (a) A técnica CMS faz a adaptação no espaço dos modelos.
 - (b) A JND da localização das formantes é maior do que a da largura de banda das formantes.
 - (c) O espaçamento típico entre tramas em sistemas de reconhecimento baseados em HMMs é de 30 ms.
 - (d) Os modelos de misturas ligadas partilham o mesmo conjunto de distribuições gaussianas com diferentes variâncias.
 - (e) As palavras para as quais o sistema de ASR fornece uma baixa medida de confiança podem ser usadas para retrainar modelos acústicos.
 - (f) O método DTW pode ser usado para reconhecimento do orador em modo independente do texto.
 - (g) Um UBM é tipicamente treinado com fala de um número restrito de oradores seleccionados.
 - (h) O algoritmo de Viterbi é síncrono no tempo.
2. Como se designam as características baseadas na concatenação de tramas de deltas?
3. Indique dois métodos de alisamento não usados neste teste.
4. Indique os dois tipos de fontes de conhecimento mais frequentemente explorados em modelos de reconhecimento da língua.
5. Qual a taxa de OOV de um sistema de reconhecimento de noticiários televisivos em Português ou Inglês com um vocabulário de cerca de 100k palavras?
 - (a) $OOV < 5\%$
 - (b) $5\% <$
 - (c) $OOV > 10\%$
6. O exemplo seguinte foi adaptado da Wikipedia.

Considere dois amigos, a Alice e o Bruno, que vivem muito longe um do outro e conversam todos os dias ao telefone sobre o que fizeram nesse dia. A Alice só se interessa por três actividades: passear no jardim, ir às compras, e fazer limpezas em casa. A escolha do que fazer num dado dia depende exclusivamente do tempo que fizer nesse dia. O Bruno não sabe como é que está o tempo no sítio onde a Alice vive, mas tem uma ideia da tendência genérica. Baseando-se no que a Alice lhe diz, o Bruno tenta adivinhar como terá estado o tempo.

Ele acredita que o tempo é uma cadeia discreta de Markov. Há dois estados, *Chuva* e *Sol*, mas não podem ser observados directamente, dada a distância (são portanto escondidos). Dado que a Alice conta ao Bruno o que fez em cada dia, *Passeio*, *Compras* e *Limpezas* são as observações deste HMM.

O Bruno conhece as tendências do tempo na zona da Alice e o que ela gosta de fazer, ou seja, conhece os parâmetros do HMM, representados esquematicamente: estados = (*Chuva*, *Sol*)

observações = (Passeio, Compras, Limpezas)

prob_inicial = {Chuva: 0.6, Sol: 0.4}

prob_trans = {
Chuva : {Chuva: 0.7, Sol: 0.3},
Sol : {Chuva: 0.4, Sol: 0.6},
}

prob_emiss = {
Chuva : {Passeio: 0.1, Compras: 0.4, Limpezas: 0.5},
Sol : {Passeio: 0.6, Compras: 0.3, Limpezas: 0.1},
}

Qual é a probabilidade da Alice ir passear no segundo dia?

7. Considere um sistema de reconhecimento de fala contínua e vocabulário extenso, independente do orador. Para um dado segmento de fala, a transcrição manual feita foi a seguinte:

Nós vemos este debate como um debate importante, mas temos grande receio, que não se passe das palavras ou seja, que haja mais uma vez uma declaração bonita, mas sem qualquer conteúdo concreto. O tratado orçamental aa é precisamente a antítese daquilo que nós precisávamos neste momento, o que nós precisamos é de pôr a economia a crescer, criar emprego, reforçar a protecção social dos desempregados e aquilo que o tratado orçamental aa apresenta é a acentuação de mais sacrifícios.

A transcrição automática produzida pelo reconhecedor foi:

Nós vemos as debate, como debate importante, mas temos grande receio. Nós pastas palavras de ou seja, que haja mais uma vez, uma declaração bonita, mas sem qualquer conteúdo compra do tratado orçamental. É precisamente a antítese daquilo que nós precisávamos, neste momento, nós pisamos a pôr a economia a crescer criar emprego, reforçar a protecção social dos desempregados e aquilo que o tratado orçamental apresenta e acentuação de mais sacrifícios.

Ignorando a pontuação e capitalização, complete os valores de H ("correct"), D ("deletions"), S ("substitutions"), I ("insertions"), N ("total"), %Corr, %Acc e %WER correspondentes.

Trata-se de fala preparada ou espontânea?

Um dos erros do reconhecedor foi causado por uma pronúncia não contemplada no léxico (/s/ pronunciado como /S/). Em que palavra ocorreu?

8. Considere o corpus de treino composto pelas seguintes frases:

O primeiro jogo junta a Holanda e a Dinamarca. O jogo é na Ucrânia. Depois Portugal joga com a Alemanha. A Dinamarca joga com Portugal. O último jogo junta a Holanda e Portugal.

Considere a frase de teste:

O jogo junta a Alemanha e a Ucrânia.

- Calcule o número de unigramas, bigramas e trigramas do corpus de treino, e a dimensão do vocabulário.
- Calcule a probabilidade da frase de teste usando um modelo de bigramas sem alisamento.
- Calcule a probabilidade da frase de teste usando um modelo de bigramas com alisamento do tipo add-one.
- Considere que existe uma classe semântica *país* constituída pelos 15 países do género feminino participantes no EURO 2012. Qual a probabilidade da frase de teste de acordo com um modelo de bigramas em que a ocorrência de cada membro da classe foi substituída pela palavra *país*?
- Existe algum quadrigrama no corpus de treino com mais que uma ocorrência? Se sim, indique-o.

(Nota: Indique as fracções correspondentes a cada bigrama, sem calcular o produto final.)

9. Desenhe a topologia típica de um modelo HMM para trifones (na última página).
10. Considere a frase *Percebe mesmo o que eu estou a dizer?* Faça a sua transcrição fonética larga e estreita, de acordo com a locução dita pelo docente. Indique para a primeira palavra quais as sílabas e os constituintes das sílabas, e quais destes são mais facilmente descartáveis. Resolva também na última página.

Respostas

Nome:	
Número:	

1. (2,4 val.)

a	b	c	d	e	f	g	h

2 a 6 (1 / 1 / 1 / 1 / 2 valores)

2	
3	
4	
5	
6	

7. (4,0 val.)

H	D	S	I	N	% Corr	% Acc	% WER	Prep./Espont.	Palavra /s/ /S/

8. (4,6 val.)

a)	
b)	
c)	
d)	
e)	

9. (1 val.) Na página seguinte

10. (2 val.) Na página seguinte